Congratulations! You passed!

Grade received 100%

Latest Submission Grade 100%

To pass 80% or higher

Go to next item

1. Which notation would you use to denote the 3rd layer's activations when the input is the 7th example from the 8th minibatch?

1 / 1 point

- ○ $a^{[3]\{7\}(8)}$
- ○ $a^{[8]\{3\}(7)}$
- ◉ $a^{[3]\{8\}(7)}$
- ○ $a^{[8]\{7\}(3)}$

Expand

Correct

2. Suppose you don't face any memory-related problems. Which of the following make more use of vectorization.

1 / 1 point

- ◉ Batch Gradient Descent
- ○ Mini-Batch Gradient Descent with mini-batch size $m/2$.
- ○ Stochastic Gradient Descent, Batch Gradient Descent, and Mini-Batch Gradient Descent all make equal use of vectorization.
- ○ Stochastic Gradient Descent

Expand

Correct
Yes. If no memory problem is faced, batch gradient descent processes all of the training set in one pass, maximizing the use of vectorization.

3. Which of the following is true about batch gradient descent?

1 / 1 point

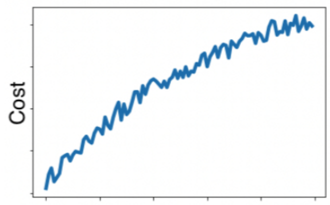
- ○ It is the same as stochastic gradient descent, but we don't use random elements.
- ◉ It is the same as the mini-batch gradient descent when the mini-batch size is the same as the size of the training set.
- ○ It has as many mini-batches as examples in the training set.

Expand

Correct
Correct. When using batch gradient descent there is only one mini-batch thus it is equivalent to batch gradient descent.

4. While using mini-batch gradient descent with a batch size larger than 1 but less than m the plot of the cost function J looks like this:

1 / 1 point



Which of the following do you agree with?

- ◉ No matter if using mini-batch gradient descent or batch gradient descent something is wrong.

- ○ If you are using batch gradient descent, this looks acceptable. But if you're using mini-batch gradient descent, something is wrong.
- ○ If you are using mini-batch gradient descent or batch gradient descent this looks acceptable.
- ○ If you are using mini-batch gradient descent, this looks acceptable. But if you're using batch gradient descent, something is wrong.

Expand

Correct
Yes. The cost is larger than when the process started, this is not right at all.

5. Suppose the temperature in Casablanca over the first two days of March are the following: **1 / 1 point**

March 1st: $\theta_1 = 10^\circ \text{ C}$

March 2nd: $\theta_2 = 25^\circ \text{ C}$

Say you use an exponentially weighted average with $\beta = 0.5$ to track the temperature: $v_0 = 0$, $v_t = \beta v_{t-1} + (1 - \beta)\, \theta_t$. If v_2 is the value computed after day 2 without bias correction, and $v_2^{\text{corrected}}$ is the value you compute with bias correction. What are these values?

- ○ $v_2 = 15$, $v_2^{\text{corrected}} = 15$.
- ○ $v_2 = 20$, $v_2^{\text{corrected}} = 15$.
- ◉ $v_2 = 15$

  ,

  corrected

Expand

Correct
Correct. $v_2 = \beta v_{t-1} + (1 - \beta)\, \theta_t$ thus $v_1 = 5, v_2 = 15$. Using the bias correction $\frac{v_t}{1-\beta^t}$ we get $\frac{15}{1-(0.5)^2} = 20$.

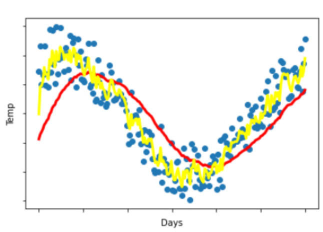
6. Which of the following is true about learning rate decay? **1 / 1 point**

- ○ It helps to reduce the variance of a model.
- ○ We use it to increase the size of the steps taken in each mini-batch iteration.
- ○ The intuition behind it is that for later epochs our parameters are closer to a minimum thus it is more convenient to take larger steps to accelerate the convergence.
- ◉ The intuition behind it is that for later epochs our parameters are closer to a minimum thus it is more convenient to take smaller steps to prevent large oscillations.

Expand

Correct
Correct. Reducing the learning rate with time reduces the oscillation around a minimum.

7. You use an exponentially weighted average on the London temperature dataset. You use the following to track the temperature: $v_t = \beta v_{t-1} + (1 - \beta)\theta_t$. The yellow and red lines were computed using values $beta_1$ and $beta_2$ respectively. Which of the following are true? **1 / 1 point**



- ◉ $\beta_1 < \beta_2$.
- ○ $\beta_1 > \beta_2$.
- ○ $\beta_1 = 0$

  ,

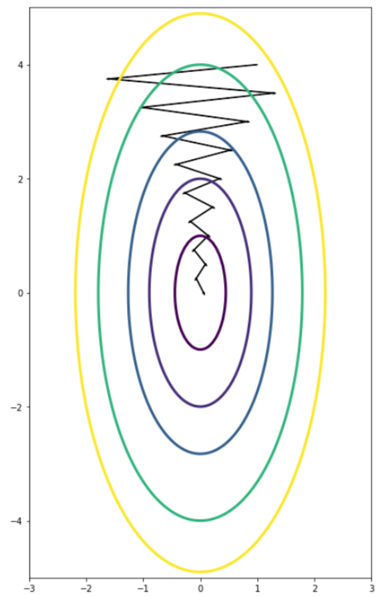
  $\beta_2 > 0$

Expand

Correct
Correct. $\beta_1 < \beta_2$ since the yellow curve is noisier.

8. Consider the figure:

1 / 1 point



Suppose this plot was generated with gradient descent with momentum $\beta = 0.01$. What happens if we increase the value of β to 0.1?

- ○ The gradient descent process starts moving more in the horizontal direction and less in the vertical.
- ◉ The gradient descent process moves less in the horizontal direction and more in the vertical direction.
- ○ The gradient descent process starts oscillating in the vertical direction.
- ○ The gradient descent process moves more in the horizontal and the vertical axis.

Expand

Correct
Yes. The use of a greater value of β causes a more efficient process thus reducing the oscillation in the horizontal direction and moving the steps more in the vertical direction.

9. Suppose batch gradient descent in a deep network is taking excessively long to find a value of the parameters that achieves a small value for the cost function $\mathcal{J}(W^{[1]}, b^{[1]}, ..., W^{[L]}, b^{[L]})$. Which of the following techniques could help find parameter values that attain a small value for $\mathcal{J}$? (Check all that apply)

1 / 1 point

- [x] Try mini-batch gradient descent.

  ✓ **Correct**
  Yes. Mini-batch gradient descent is faster than batch gradient descent.

- [x] Try using Adam.

  ✓ **Correct**
  Yes. Adam combines the advantages of other methods to accelerate the convergence of the gradient descent.

- [ ] Try initializing the weight at zero.
- [x] Normalize the input data.

  ✓ **Correct**
  Yes. In some cases, if the scale of the features is very different, normalizing the input data will speed up the training process.

Expand

Correct
Great, you got all the right answers.

10. Which of the following statements about Adam is ***False***?

1 / 1 point

- Adam combines the advantages of RMSProp and momentum
- The learning rate hyperparameter α in Adam usually needs to be tuned.
- We usually use "default" values for the hyperparameters

β_1, β_2

β_1, β_2 and

ε

Expand

Correct